

2.1: Displaying Data!

- Organizes data
- Helps show similarities and differences among different groups
- Shows **distribution** of the variable.

- picture of the data
- What values the variable takes (D, R, E)
- How often it takes the different variables (# or %)

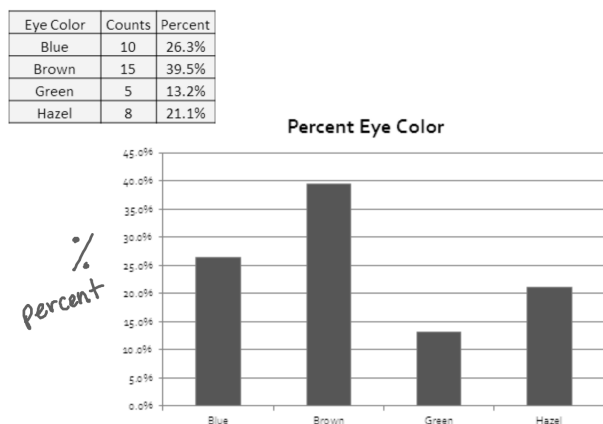
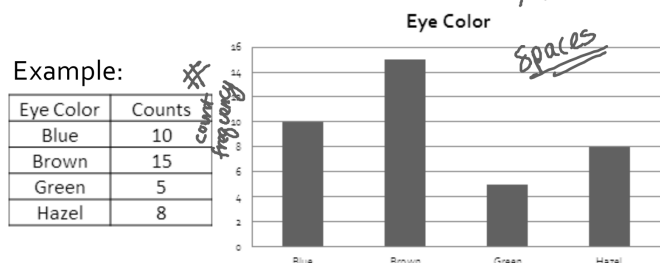
Examples: bar charts, pie charts, histograms, dotplots, stemplots, boxplots, etc. *Quant.*

2.1: CATEGORICAL DISTRIBUTIONS:

Bar Chart:

* Vertical or horizontal bars that represent the **counts** (frequency) for each value of the variable.

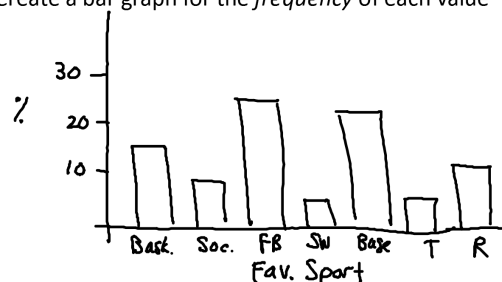
* Can also show % (relative frequency)



Example: The distribution of favorite sport is as follows:

Basketball	12	15.4% x	Baseball	18	23.1% x
Soccer	8	10.3% x	Tennis	6	7.7%
Football	20	25.6% x	Running	10	12.8%
Swimming	4	5.1%			

Create a bar graph for the **frequency** of each value

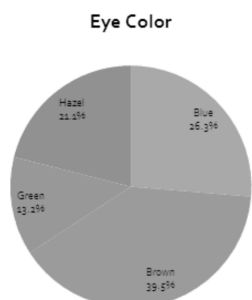


Pie Chart:

- A circle represents the whole and each value of the variable is a piece of the circle.
- Can only be used when comparing parts to the whole.
- Always in % (relative frequency)

Example:

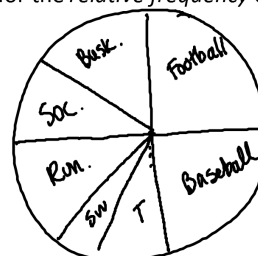
Eye Color	Counts	Percent
Blue	10	26.3%
Brown	15	39.5%
Green	5	13.2%
Hazel	8	21.1%



Example: The distribution of favorite sport is as follows:

Basketball	12	Baseball	18
Soccer	8	Tennis	6
Football	20	Running	10
Swimming	4		

Create a pie chart for the **relative frequency** of each value



Example of something that CANNOT be a Pie Chart:

Turn to page 39 in the textbook (Example 2.3)

Why can this NOT be a pie chart??



Examples:

Using the class data that you have, create the following:

- 1) A bar chart of "Hair color"
- 2) A pie chart of "Favorite Subject"

TWO WAY TABLES:

- Used to compare observations for two different categorical variables
- Example** – 200 adults at a supermarket were asked for their educational level and whether or not they smoked.

		Education		
		High School	2 yr college	4+ yr college
Smoke	Smoker	32	5	13
	Non-Smoker	61	17	72

We always want to find the totals:

	High School	2 yr college	4+ yr college	Totals
Smoker	32	5	13	50
Non-Smoker	61	17	72	150
Totals	93	22	85	200

Marginal distribution: The totals, converted to %.

	High School	2 yr college	4+ yr college	Totals
Smoker	32	5	13	50
Non-Smoker	61	17	72	150
Totals	93	22	85	200

Smoker:

Yes $\frac{50}{200} = 25\%$

No $\frac{150}{200} = 75\%$

Education:

HS $\frac{93}{200} = 46.5\%$

2yr $\frac{22}{200} = 11\%$

4+ $\frac{85}{200} = 42.5\%$

Example: Hair color vs. Gender

	Brown	Blonde	Black	Red	Total
MALE	26	24	10	3	63
FEMALE	20	35	12	6	73
TOTALs	46	59	22	9	136

- 1- What are the 2 variables? Hair Color & Gender
- 2- What are the values of each of the 2 variables?
- 3- What is the marginal distribution of the ROW variable?
- 4- What is the marginal distribution of the COLUMN variable? (2) Hair Color: Brown, Blonde, Black, Red
Gender: Male, Female

③ Gender

$$\text{Male} = 63/136 = 46.3\%$$

$$\text{Female} = 73/136 = 53.7\%$$

④ Hair Color

$$\text{Brown} = 46/136 = 33.8\%$$

$$\text{Blonde} = 59/136 = 43.4\%$$

$$\text{Black} = 22/136 = 16.2\%$$

$$\text{Red} = 9/136 = 6.6\%$$

Questions: Answer the following questions

	High School	2 yr college	4+ yr college	
Smoker	32	5	13	50
Non-Smoker	61	17	72	150
	93	22	85	200

1- What percent of people are smokers? $50/200 = 25\%$

2- What percent of the people had 4+ years of education and were smokers? $13/200 = 6.5\%$

3- What percent of the smokers had 4+ years of education? $13/50 = 26\%$

4- What percent of those with 4+ years of education were smokers? $13/85 = 15.3\%$

	High School	2 yr college	4+ yr college	
Smoker	32	5	13	
Non-Smoker	61	17	72	150

5- What percent of the shoppers that were non-smokers had 4+ years of education? $72/150 = 48\%$

6- What percent of the non-smokers had 4+ years of education? $72/150 = 48\%$

7- What percent of shoppers had 4+ years of education given they were non-smokers? $72/150 = 48\%$

	High School	2 yr college	4+ yr college	
Smoker	32	5	13	
Non-Smoker	61	17	72	150
	93	22	85	200

1- What percent of the shoppers were non-smokers? $150/200 = 75\%$

2- What percent of the shoppers had a high school education only and were non-smokers? $61/200 = 30.5\%$

3- What percent of shoppers were non-smokers given they had a high school education only? $61/93 = 65.6\%$

4- What percent of non-smokers have a 2 year college degree? $17/150 = 11.3\%$

5- What percent of shoppers with a high school degree only are smokers? $32/93 = 34.4\%$

6- What percent of those with a 2 year college degree were smokers? $5/22 = 22.7\%$

Answers:

1- 75%

2- $61/200 = 30.5\%$

3- $61/93 = 65.6\%$

4- $17/150 = 11.3\%$

5- $32/93 = 34.4\%$

6- $5/22 = 22.7\%$

Try the worksheet 2.1A: Categorical Practice

2.1 A]

① Row Var = AGE

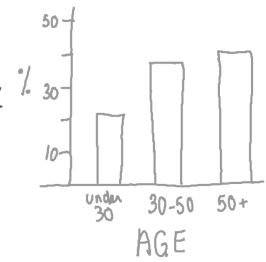
② Column Var = POLITICAL PREFERENCE

③ Age

under 30: $296/1442 = 20.5\%$

30-50: $560/1442 = 38.8\%$

50+: $586/1442 = 40.6\%$

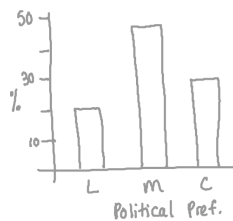


④ Political

Liberal = $290/1442 = 20.1\%$

Mod = $704/1442 = 48.8\%$

cons = $448/1442 = 31.1\%$



⑤ $290/1442 = 20.1\%$

⑧ $88/586 = 15.0\%$

⑥ $83/290 = 28.6\%$

⑨ $280/704 = 39.8\%$

⑦ $586/1442 = 40.6\%$

⑩ $140/1442 = 9.7\%$

⑪ $284/704 = 40.3\%$

⑫ $161/1442 = 11.2\%$